Jon Krosnick: Okay. And now for something not completely different, Matt Berent is going to pick up on this theme. Matt has been working with the American National Election Studies for a number of years now, and was the lead on the project that Michael has alluded to, where we were looking at the relationship of voters, survey respondents reports of turnout with what could be done with government records that Matt will describe.

Matthew Berent: Well, I want to thank Stanford and the National Science Foundation for inviting me to come here and talk a little bit about some work I've done with Jon and Skip using official records or government records, the kind of records that Mike was just talking about, to correct supposed errors in turnout self-reports.

Now, the way surveys have been discussed in the news recently, you might think the only reason for conducting a survey is to predict what's going to happen on Tuesday. Well, it might be strange, then, that we're asking a question about what somebody did in the past. Did you vote?

First of all, there's nothing to predict. They've already done it. And secondly, if you really want to know how many people voted, why don't you just go to the government statistics? To statistics that every state publishes about who voted?

Well, the reason is the goal is not to make a prediction about who voted. The goal is to understand why people vote. It's about participation. And the only way we can understand voting, why people are turning out, is to link turnout data – accurate turnout data – with other things we know about a person. Then we can understand why a person goes to the polls, what are the facts? And we're in a position possibly, to increase political participation in the country.

Now, in order for us to that, we have to link accurate turnout data with other data. And we're going to make some assumptions. First we have to assume, if we're going to make statements about an entire population, we have to assume that we've effectively sampled from that population, we've designed a survey that can effectively produce the type of data we're interested in, those data are accurate. And from that, we can make – we can develop and test – models of the population. We can understand what’s going on.

Now, a Dutch mathematician once said, “In theory, there's no difference between theory and practice. But, in practice, there is.”
We take this theory of how the survey should work, and we find an apparent disconnect. If you look at what the American national studies – election studies – have found since the 1960s, the percent of individuals in the survey who report that they turned out is consistently higher than the percent you would expect, based on the official statistics published by the states.

And that discrepancy is often 20 percent, or around 20 percent. Now, what's encouraging is that those two lines kind of track together. When actual turnout goes up, self-reported turnout goes up a little bit. So, there seems to be some accuracy, but this discrepancy between what people say and what you would expect is cause for potential concern.

This leads people to conclude that estimates from our samples do not reflect the population because they don't match what we know to be true based on the government statistics. And there's a couple of reasons for these discrepancies. If you ask a person, “Did you vote?” Simple question.

Some people will not answer the question. That's a non-response. Either they decline to participate in the survey, or they declined to answer that particular question, so we don't have data for those individuals. It's not that these people are giving us bad information, they're just not giving us any information at all.

There are survey effects. Possibly participating in a survey influences the behavior in which we're interested in. The studies in the graph that I just showed you, those all involved a very in-depth interview before the election, and then an interview after the election. And something about answering all those questions about politics before an election might change the behavior – the voting behavior – of individuals. So, participation affects behavior.

But, another possibility, and Jon spoke about this in his introduction, is there's misreporting. People are giving you the wrong answer. They give you information about turnout that's not factually correct.

Now there's different types of misreporting, reasons why a person might give you bad information. One is they misinterpret the question. You ask, “Did you vote?” And the respondent's trying to think, “All right. What's this person really asking me about? Well, maybe what they really want to know is if I think politics is interesting, or if it's important. Well, of course I say I think it's important. So, I'm going to answer the question in the way that I
think it should be answered.” So, they misunderstand what the question is really asking.

They might misremember. They don't know whether or not they actually turned out on election day, but they do remember that they wanted to. They talked about politics for a lot of colleagues, a week coming up to the election, so they figured, well they must have. They must have voted.

Or, a big concern is intentional lying. That is, people know what the right answer is, and they're deciding to give an answer they know is wrong. And the reason is a person might be thinking to themselves, “I didn't vote, but if I say that I didn't vote, I'm not going to look good to this person to whom I've been speaking for the last 45 minutes. So, I want to make myself look good, so I'm just going to say 'yes' even though I know it's not true.”

This intentional lying explanation for that discrepancy between the self-reports and the government statistics is an explanation that's got a lot of traction. The *Wall Street Journal* assumes, and told the world, that people lie in surveys. We all know that.

Well, what are we going to do then? We have this apparent problem. Many people believe it's true that people who are participating in our ANES surveys are lying. So, the solution is to use official records. And I want to thank Michael for all of that stuff he talked about. Where is he at? Ah. Michael. For all of the stuff he talked about leading up to this.

And what we attempted to do was use the type of records that Mike was talking about to determine what an individual's actual turnout was in our surveys. So, in order to do this, what we had to do is first obtain those government records. That included both who was registered to vote, and their turnout histories.

We had to find some strategy for matching each of our respondents to his or her record – government record – and once we did that, we could figure out what the correct – we didn't have to rely on what the survey respondent said. We could just look at what the government record said.

If we follow this process, we can identify, presumably, people who are lying in our surveys. And we do this two ways. One, a respondent said one thing about turnout. They said, “Yes. I voted.” Well, in the process of matching, we find that person's record, and the government record actually shows this person did
not turn out on election day. So, that's a pretty strong case that this person has not told the truth. The government records effectively discredits the self-report.

But another use, another way to identify lying is when you don't find a record. Now, talked a lot about the quality of the records and the matching, trying to match survey respondents to government records. Well, if an individual sent says they turned out to vote, but they're not registered to vote, then there's no record to be found for them, because the records we obtained are only for people who have been registered in a state.

So, if an individual says they turned out, be we can't find a record that that person is registered, then we have conclude the person isn't registered, did not turn out, and therefore is lying or has lied in the survey. And that's what we call the fail to match problem. That is, we have to make inferences about the respondents for whom we can't locate government records.

Now, we wanted to do this with the ANES panel study from 2008, 2009. And we have a fairly good sample, effective sampling, random digit dialing, our population of US citizens, voting age, land line households. What's different about this survey than the ANES surveys that I talked about before is that those were all in person interviews. One conducted before. One conducted after.

The difference here – one of the differences – is that these respondents all completed monthly online surveys rather than the face to face interviews. And they completed these over a number of months.

Included in the November online survey was this question, which was the attempt to measure voter turnout, and there's one option, “Did not vote,” a number of options for voting, and another for people who weren't sure, who couldn't recall, and if individuals selected that, we asked them a follow up. “Do you think you probably voted or probably didn't vote?”

So we were giving people a lot of opportunities to answer the turnout question in different ways, really prompting their memory – [coughs] excuse me – reducing some of those memory biases. So, what happened in 2008 with these panel study respondents, compared to the time series respondents?

We basically found the same thing that we've been finding for years. More people said they turned out to vote than you would
expect based on the number of people who actually turned out in 2008. And the difference was striking, 24 percent. Now this is well beyond the margin of error for a survey. We've got a discrepancy of 24 percent. So that gives us some cause for concern that people maybe, perhaps, our data are inaccurate.

So, the task was to determine whether or not our turnout data were accurate, which is the validation where we try to match government records to our survey respondents and determine the correct turnout for each of the respondents. But there's two big problems for us at ANES. One is there are more than 200 million people who are eligible to vote in 2008. That's a lot of data to sift through – potential data – to try to locate records for 400,000 respondents.

And some records, as Mike indicated, are very difficult to obtain. Virginia doesn't make its records available. States like Wisconsin are available if you've got a lot of money. It's several thousands of dollars to get the records from those states. Other states, much easier.

Thankfully, there are commercial vendors who do this type of stuff. So we approached them to possibly do this work for us, and actually two reasons. One, maybe they could do it for us, but if they couldn't do the records matching for us to get the accurate, or match our survey respondents to the government records, then perhaps we could learn something from them about how they do it and we could do it ourselves.

So, we actually approached 22 different commercial vendors who do this sort of work. These companies maintain, many of them, national databases, which include the government records. As Mike indicated, they supplement these with commercial data. And they can clean the data to increase its accuracy, and all of them have some technology that they've developed that allows them to do this fairly efficiently, to go through millions and hundreds of millions of records, and match those records to other data.

But, in order for us, as the American National Election Studies, to use one of these vendors, given that our data, or those data, are made available to any scholar who wants to use them, there's an obligation that we provide full disclosure about where the data come from, how they've been developed, how they've – basically how they, what is its accuracy?
So, when we talk to the vendors we said, “We'd like you to look at doing this, but in order for us to let you do it, or have you do it, we need to know when your records were obtained.” Because, as Mike indicated, states differ in how they maintain records, and even within a state, the time – the timing in which records are updated – differs wildly.

In Florida, some jurisdictions update their turnout records shortly after an election. So, if a person voted in November, in December their government record indicates that they voted. But other jurisdictions, it might be 12 months later before that record is updated. So, we needed to know when the records were obtained.

We needed to know from whom they were obtained. Are they getting records – their records – right from the state? Or are they going through a third party? And if they're going through a third party, where did the third party get it from? And what did they do to the data before they gave it to you?

And we also needed to know the process by which they were cleaning their data. How did they decide if they're combining two records, which piece of information is right? How do they first match those two records, and how do they alter the data to make it more valid?

And we also needed them to fully disclose the matching process. And this is the black box. Now, every company said, “We can do this.” And some companies provided us some information. But, for the matching, we needed to know, what’s the process? What's the decisions?

What information do you use to make these decisions about matching, and how do you determine whether you have matched a record or have not? And that's the issue of coverage, the high probability versus low probability. What are your decision rules?

Now what we learned in talking to these vendors is that basically all of them wanted three types of information in order to do the matching: Names, addresses, and dates of birth. If we had all that information is in all the government records. So, that was the starting point. Needed that information. We also learned that some vendors are more transparent than others.

And I think, Tom, you were alluding to the fact that the more you engage them, the more they'll give you – the more information they'll give you. In talking to one vendor, they sent me
spreadsheets that indicated the date on which they received the file from each state for their government records. They also indicated here's the fields, here's the four different combinations that we incorporate into our algorithm.

So, some groups were very transparent up to a point. Others were simply, I asked a question, and they said, “I don't have time for this. So, I'd be happy to send your business somewhere else, because I don't want to talk about our process.”

Some vendors do match at these different levels of confidence, the high probability to low probability. But, no vendor was going to give us insight into the black box. Nobody was going to say, “Here's the program that we use.”

And this makes sense. This is their secret sauce. This is what gives them a competitive advantage. They don't want to – if they tell us, and we disclose that to everybody, then all of a sudden there's no need for Catalyst anymore, or Aristotle, or Labels and Lists, or any of the 22 companies that we talked to.

But because nobody was willing to be fully transparent about the process, we, based on our obligations to the users of our data, could not use them. We had to do this ourselves.

So we had to obtain the government registration and turnout records from a sample of states. Now, we had – there's no way we could afford data from all of the states that were willing to make it available to us, so we had to limit it to a sample. And, we had to find some application that allowed us to sift through these millions of records to do the matching for us.

This is the program that the vendors use, but they won't tell us what it is. They won't – and I'm not implying that this program that I'm going to give you the name of is something that the vendors use. We had to find something comparable to what the vendors might use to do this matching for us.

And we did find there are a couple available and one that's publicly available. It's called LYNX Plus, developed by the CDC. So, anybody interested in doing this sort of matching themselves, rather than relying on a vendor can download this application and give it a try.

So, we had to use a subsample of data, and for this, we selected respondents in our survey who lived in one of these six states.
You're thinking, “Why these six?” And there's actually three reasons. The first is we had pretty good numbers of respondents in these different states. Two, in talking to Mike, these states do a fairly decent job of keeping good records. And three, actually the more important ones, these were things, information, records we could get fairly easily.

Pennsylvania, as a resident of Pennsylvania, I was able to contact the state and say, “I want the voter lists.” They provided them. California, I filled out a form, sent in a check for $20.00, they gave me a CD with all of their data shortly after 2008. So, these were easy to obtain and in our work, we're actually weighting everything, so we're reflecting what's going on within different states.

Now, talking about coverage and the probability of matching individuals, we followed that lead and developed three different levels of accuracy, or three different methods of matching. For a high probability of a survey respondent being matched to the correct government record, we had specified that the names, addresses, and dates of birth provided by a respondent to us had to be absolutely identical, character by character, to the records – to a record – that we had in a government list. Only then would we say we've found the right record for this person.

But, Mike also noted there's tons of errors in this data, and that's going to allow us – that's going to result in us – missing some cases. So, we developed a different level, probability of accuracy we called “moderate.” Individuals had the same or similar first and last name, same or similar address.

So, these could be differences that could be explained by a typographical error, a transposition of letters, a missing character, or perhaps, an informal name like “Pat,” and a formal name, “Patricia,” in one record and the other. This allowed us to presumably address some of the things that the commercial vendors do in correcting for discrepancies in records.

And for our least strict, which had the lower probability of a match, but still some likelihood that we've located the proper records, names, addresses, and dates of birth had to be same or similar. We also allowed records to be matched if we had – people had – the same or similar first and last names, and an identical date of birth, to try to capture those people who may have moved between the time the government record was created and the time that we did our survey.
But also, there's some people who might provide a very informal name like “Butch” to us, and we can't tie that to a particular formal first name. We didn't want to miss all of those individuals. So, our third level, our third type of match in this least grouping, a record had to have the same, or the identical last name, address, and date of birth, but could have a completely different first name.

So, we had different levels of matching, corresponding to different probabilities of correctly identifying the record, or coverage, as some of the vendors have called it.

Now, what did we find? Well, when we used this, which was presumably very comparable to what commercial vendors would do, we're assuming, according to the strict matching method, according to records that we've matched to our respondents using this strict set of criteria, only 43 percent of them actually turned out. Quite a low number.

The moderate, that method suggests 59 percent turned out, and the least suggested 69 percent turned out. Now, when we asked our respondents how many of them, or if they turned out, 83 percent said yes. According to the validation strategy, anywhere from 43 to 69 percent turned out. Quite low.

Now, in the population, around 62 percent turned out in 2008. Now, if our criteria for evaluating the validity of this method of taking government records and using them to learn about our respondents or validate what they say, if the correspondence, or if the validity of that method is based only on the similarity between the population and the values that you get based on this matching, we'd have to conclude that the moderate strategy is the best. We're missing by only three percent. Okay? So, not too bad.

So this would suggest if lying in the surveys is the reason for this discrepancy, this method of matching actually does a pretty good job. We're getting up here, apparently, more validated when we use this moderate strategy. Least matching is sort of in the ballpark. Not too bad. But this strict, high coverage, you need a really high probability that you've located the proper record for this individual doesn't seem to be working, because the difference is so large.

Yep.
Male 3: Are these cumulative? So if you're a least match, are you also – if you're a strict match, are you also counted –

[Crosstalk]

Matthew Berent: Yeah. Yeah. I'm sorry I should've done it.

Male 3: – as matched in the least? Okay.

Matthew Berent: If anybody who was matched to a record based on our strict criteria also qualified for the moderate and the least. So, that is cumulative.

But regardless of the matching method that we used, it would suggest that these self-reports simply can't be trusted. People appear to be lying in our surveys. And if they're lying in their surveys about that, they might be lying about other things, so why do a survey at all?

Well, you might think this strategy of using government records actually solves the problem. We have this apparent problem where respondents lie, we use the government records. We find that when we do that, we have turnout in our sample that's close to the turnout in the population, therefore, we've dealt with this.

But if this is true, then it's also the case that we should find that the sample registration, that the number of people in our sample who are matched to a registration record reflects the percent of individuals in the population who are registered. Both those rates should be comparable.

Well, according to the different states, around 83 percent are registered to vote in those states. In the population that we're interested in. When you use the strict matching method, you would conclude that only 43 percent of your sample respondents are registered to vote, a rate that's quite a bit lower. So, in this case, we've got the strict method that's actually dramatically underestimating registration rates.

Moderate, which did pretty well for turnout, now doesn't work for registration. It actually underestimates registration by about 17 percent, making it apparently invalid. And even our least strict method underestimates registration rates, in this case by around 5 percent.
So, we're trying to use government records to get valid information about whether our respondents are registered and whether they vote, and we find that strict is really bad. This if you require a high probability, you're going to get numbers that are way off.

So, when you use a moderate – what we called our moderate method, you do pretty good at turnout, but not registration, and when you use this least strict, low probability, you have turnout that's too high, but close to registration. So you think, “What's going on?”

If this government, if this procedure for matching or using government records is valid, shouldn't we be getting good numbers across the board? And we're not.

Now we think, when it comes to using government records, there are actually two biases that are going on. The first is a downward bias, that when you use government records, you're going to incorrectly identify some people as having not turned out in this case. That is, you're going to say, based on our work in trying to find your government record, we're concluding that you didn't vote even though you said you did. And that number is going to lower the rate in the sample.

Now there's also an upward bias. Something that people haven't considered as a major cause in the past is that the people who are participating in our ANES surveys are actually more likely to turn out. Suggesting that maybe the answers, the self-reports, are not invalid, they're not – people are not lying, they're actually telling the truth, and they're more likely to vote.

So, we wanted to figure out if that's the case, we wanted to provide some evidence of an upward bias. So, to do this, we actually looked at government records, but in this case, we're looking only at the people in our sample that we have matched to a government record using one of our methods. So, if we matched somebody to a government record using the strict method, we're very sure that that person is actually registered.

So we can look at all of those people, and look at the – among them – what percent actually voted, compare that to the number of people in the population who are registered who also voted. Those numbers should be fairly comparable.

Well, in the population, turnout among people registered is around 75 percent, or was 75 percent in 2008. According to our different
methods, 88 percent of the people who are registered in our sample according to this strict method of matching records actually voted. Moderate, a few more. Least, about the same. It's fairly consistent across those.

Now, regardless of the method used to match the respondents to the government records, they're all showing that our survey respondents are more likely to turn out than people in the comparable population.

So, is there any evidence of a downward bias? That is that we're making incorrect conclusions when we use government records. Well, for this, we want to look at individuals who said they turned out. So, their self-report is, “Yes, I voted.” Among these individuals, when we use government records, your determination of how many of them are telling the truth will depend on the type of decisions you make, or the type of matching you do.

You'll determine that a ton of people in your survey are lying if you have a very strict matching rule. As you relax the criteria, you find more and more, “Hey, it seems that our respondents are telling the truth.” So, someone we conclude is lying based on our strict method, we would conclude is actually telling the truth based on our least strict matching method.

Now, the number of individuals where you can make a legitimate case that it was beneficial to locate these government records, or to use the government records is a small number here. These are people who they say they turned out. We find a government record for them, and the government record says, “You didn't turn out.” So, we've got a pretty good case that they did not tell the truth. But that number is very small.

The overwhelming number of individuals whom we identified as not telling the truth are based on this fail to match. Individuals where we conclude we can't find any record of your turnouts, therefore you must not have turned out. We can't find a registration record for you.

Now, I think in order to move forward using government records to try to validate what people say about turnout, we need to know whether decisions are made on a failure to match or matching to a record that doesn't show turnout. And I think that's an important consideration. And addressing the fail to match problem. What do you do about the ambiguity cost, by your failure to locate records for a person who's participated in your survey?
So, we started off with this supposed problem that respondents lie. And we based that, our conclusion that this is a problem, based on the difference between turnout in our sample and turnout in the population. But it turns out that respondent lying is not really as big of a problem as we thought. The problem is that turnout in our sample is actually bigger than turnout in the population.

So, I think one of the possible avenue for more work is to determine how do we keep our sample if we've recruited them efficiently, we're asking good questions, how do we make sure that the process of interviewing them and surveying them doesn't alter the behavior in which we're interested in?

Some people originally have, and you'll hear people talking, big advocates, proponents of the use of government records, to overcome the problem of respondent lying. Now, I'll show that respondent lying is not as big of a problem as we thought it was. But still, using government records holds some opportunities for us. There's additional information available in those records that we could link to our survey respondents.

But in order for us – so, using government records to solve the problem of respondent lying is really not a consideration. I think that using government records have additional uses that we would like to explore, and I think what we've discovered in this process is that often, using government records, this process of matching survey respondents to their government records, has these competing biases, which only creates the illusion of accuracy – an illusion based on the outcome similarity between the rates in the survey or the sample, and the rates in the population.

So, what's the solution? At least when it comes to turnout, is to use these self-reports. Don't rely – self reports appear to be, at least on turnout, incredibly accurate. But I think there are other uses for government records, but we need to explore the technology. We need full disclosure. We need to understand how that works, and explore different ways with which we might supplement information we have about our survey respondents. Okay.

*Randy Olsen:* Randy Olsen. Was the error in self-report versus administrative reports of voting systematically related to the states that were top down versus bottom up inasmuch as you could have multiple registrations for the same person for either legal reasons, or just if it's a bottom up, just error?
Matthew Berent: Yeah. I don't have information on the top down versus bottom up, but what I can tell you is we did break down all of the types of work that I presented here by individual states. And what you saw was simply aggregated across the states. When we start looking at which method of matching worked best in which state, then we start to find a lot of differences.

We find, well, in some states, the strict – in no state did the strict work really well, but in some states, our moderate criteria worked pretty good. In other states, the least method produced results that were at least comparable to the population parameters. Now, that doesn't mean that – what that tells us is that the effectiveness of the matching process varies across the state as a function of the quality of the records of those states.

So, some states have a lot of dead wood in them, more so than others. Some states, look at – you're talking about dates of birth. You'd be amazed at the percent of the population in the different states that are born January 1, 1900. It's just unbelievable. That's the placeholder. That appears in a lot of government records.

So, and some states, I wouldn't know if they're top down or bottom up, but some states have much more error in them than others, causing different types of problems in a different matching methods.

Andy Peytchev: One quick question and two comments. So – oh. Andy Peytchev. So, the question is would you speculate how the results were affected by the panel design and the survey in terms of comparing it, how replicable would these findings be to a survey where you go in and the household that you haven't interviewed before, and you ask them, “Have you voted?” As opposed to a panel that you've built rapport over time giving them surveys every month?

And then, the two general comments, I suppose more about thinking about NSF, it seems there are different areas of research that they're even broad there that would benefit research, and this would be definitely one of them. The linkage of linkage methods would be one.

I think the more sophisticated methods that have been developed by people like Bill Winkler have been more about he wants to link and have a degree of certainty that you have the right link, but it doesn't have the purpose in mind where you're looking for something that's unbiased. In other words, you're, in this purpose,
you would rather allow some error, as long as on average, you're getting it right, as opposed to having that perfect match.

And then the other would be the different databases that you could be thinking about in terms of we've been talking about administrative data and having error in them, but I suppose I think of each database having its own error properties, and its own different purposes it could be fit for use, and thinking about that and exploring each major database.

So, you could think about the uniform crime reports and the national crime survey. Well, how could they be incorporating these? Or how could you use Experian data across multiple surveys? And that's – I'm also involved on research in at least looking at things that Experian data could be useful for, for example.

But then, the broader issue for, maybe NSF will be thinking about what are these major sources of data and what are they fit for use? Is it non-responsive adjustments? Is it linking at the personal level and having a better understanding of phenomena, and so on?

Matthew Berent: Well, on the first question, the difference between what we've observed with the panel study respondents versus the time series respondents – and I think we start saying we find the same magnitude of a discrepancy between the time series respondents and the panel study respondents. That leads us to think whatever was going on with time series might be going on with, or probably is going on with the panel study respondents.

Now, we don't know that that's the case, and I think that's why, moving forward, we need to replicate this with time series respondents who have done those very in depth face to face interviews before and after to make – to see what type of – to get a better understanding of the effect that appears to be driving the changes in behavior.

But I think, based on what we have right now, I can't offer a definitive statement that the same effects would hold up. Only suspicion.

And on some of the other points, I think when it comes to incorporating this with larger databases, I think that's a wonderful idea. I think what the vendors have told me they do, which is they get the records from the state governments. They supplement
registration and turnout histories with records from the DMV, marriage licenses, who knows what.

They also then go to another party, a third party, and buy consumer information that they link up with the government records. And I think – try to get all of those, if it's possible to get records from all of those into a single database, yet protecting confidentiality of individuals who might be in that database, and using those type of data might be – yeah. The data, presumably, would be cleaner.

Now the vendors tell us that this is what they do, but we don't know how they do it. So that creates a problem for us in wanting to distribute those data to other individuals.

Matthew DeBell: Matthew DeBell, Stanford University, with an ANES at Stanford. I just wanted to say something in response to Andy's question. I think Matt said that the differences, or that the turnout, the magnitude of the errors between the face to face time series study and the panel study are comparable, and they're in the same ballpark, but I think they're a few points higher, a little bit, in the panel study.

And I think that the reasons for that probably point to sampling differences being a bigger source of error than lying. And the reason that I say that is that the – well, for one is that social desirability effect would presumably be smaller with the online mode than when you're talking to an interviewer. You're not going to be as embarrassed to tell a computer that you didn't vote, as you would be compared to telling a person that you didn't vote.

And another reason for possible differences would be the panel sample is an RDD sample subject to attrition, and subject to panel conditioning effects, and both of those would be more pronounced in the panel study than they would be in the face to face survey where there is, potentially, both attrition and conditioning, but it's from one pre-election interview, whereas with the panel study it's over many months.

And then, I had one other question – or I had a question. So, I'm going to make Gary's head explode now.

Gary Langer: Okay.

Matthew DeBell: So, brace yourself. The ANES is currently contemplating the acquisition of turnout data from these vendors. And so, we face this problem. Right? On the one hand, there's this lack of
transparency with the vendors. On the other hand, you've demonstrated that the task of going out and doing the validation ourselves is frankly impractical, because we can't get to all the states. We can't get all the data that the vendors are able to get.

So, the thinking on the part of the ANES PIs is that this is, for us, a relatively low cost proposition that we can get these data, and then what's incumbent upon us is to be transparent about the lack of transparency that's available to us. So, we want to say, “Here's the data we've got. Here are the reasons” — and I appreciate the comments by Josh and Gary about why we might be skeptical, but here are the reasons — “that we should be skeptical about these data. Now analysts, go forth and do with them what you will.”

So, but in an attempt to get a better sense of how good they are or how bad they are, my question for you is whether you think that the kind of work that you've done to do validation, what would happen if you got the data from the vendors for the same addresses that you did this work on, you compared your results to what they have to say? Do you think that you could use your results to estimate the accuracy of what the vendors provide? Or would there be minefields or problems in trying to reach conclusions of that nature from the work you've done?

Matthew Berent: No. Let me see if I understand. If you had the respondents and government records matched by one of the vendors, and you were to send me the same information that the vendors used without knowing what their black box was –

Matthew DeBell: Mm-hmm.

Matthew Berent: – would I be able to make inferences about not necessarily what's in the black box, but the accuracy of their data?

Matthew DeBell: Yeah.

Matthew Berent: Um...

Matthew DeBell: Yeah. I'm not trying to reverse engineer their process. I'm just trying to look at their results and saying, “Is this good or is it junk?”

Matthew Berent: I think [laughter] – well that – I don't want to set myself up as a standard for evaluating what's junk or what's not. I think there are other information that you can get from vendors that will allow you to evaluate their performance. I think when you look at some
of the work that's been published that's used a commercial vendor, you find turnout rates based on the commercial vendor that's very comparable to the population turnout rates. But what you don't often get is the comparison between population registration and the sample registration.

Now, in one study, in which the commercial vendor's results were published, the results for turnout were pretty much the same, but their estimate of the number of people who were registered based on matching was about 20 percent lower than registration in the population. So, when I see that, I think, “Aha.”

And I didn't actually have to do anything with their database. I just had to look at their results and say, “You've got one number here that's a lot lower.” And if there's one thing that should be unaffected by the – it's possible that when you participate in a survey about politics, you're more likely to vote.

And it could be you're more likely to go out and register to vote. It's difficult to imagine participating in a survey, and as a result of that, you're more likely to go to your local jurisdiction and say, “Please take my name off your registration rolls.”

So, if I find results from, as I've seen from one of the commercial vendors, where their work suggests a lot – a very low proportion – of individuals are registered, that sets up a red flag. That tells me they're having a big – there's a big failure to match problem. They're not finding records for individuals who are registered, and they're making conclusions about their registration that are wrong.

Male 4: I just had a quick answer also, which would be I think I take the one implication of Matt's work is that if you want to evaluate whether Catalyst is accurate in figuring out whether a set of survey respondents voted or not, just compare it to what the respondents said, because the respondents are right. And so, if Catalyst doesn't match them, they're wrong.

Response: [Laughter]

Gary Langer: If I could, it's Gary Langer. Just one other question. It seems to me to some extent that this discussion is based on the assumption that the sample's good, that we have an unbiased sample of respondents, and that's why we expect there to be a match between the self-reported behavior and actual behavior.
But it also seems to me that we know that our samples are not always good, or certainly not always perfect, and I wonder, for example, about social engagement bias. In other words, the people who are more likely to do socially engaged things: Participate in community activities, give to charity, attend church, vote. Are the same sorts of people who are disproportionately likely to participate in surveys?

And we get more people saying they voted in surveys not because they're lying, not because of a mismatch in record, but because it's true. Possible?

*Matthew Berent:* Yeah. I think there's some, there's a possibility there, and I think there are ways to look at that. What we presented was a possibility that somehow participating in the survey causes people to change their behavior. But what you're suggesting is that people who have different types of behavior are actually more likely to participate in the survey.

And I think for that you have to evaluate the sampling procedure. And I think ANES, more than other organizations, really does a good job at trying to recruit people that are representative of the population. But, it's possible that those methods are – could be – improved.

*Male 5:* Good. No, I had the same idea about this idea of the sample and the issue of people who participate in surveys and it came from the thought about the ANES panel. I know a little bit about panels, so I was – because my concern, we did a study one time for a group that was looking at generosity in terms of people contributing funds and money to various causes, and so on, and so forth.

And they, this particular client, asked us to look at, also, the rates of volunteerism on that panel and compared it to census data on volunteerism. And we found that the panel members that responded had higher rates of volunteerism than the general population. That made us think very seriously about the possibility of adding volunteerism using weighting statistics, weighting to the CPS data, to modify the panel.

And their conclusion was the generosity levels of the population would be more accurate when we modified the data to make it more closely reflect benchmarks of volunteerism in the larger population. So, therefore, do people who volunteer more join panels? Do people who volunteer more, for example, do surveys and agree and are they more likely to vote?
And that, you know, so I guess issues about that with regard to people who participate in surveys and their voting propensity need to be taken into consideration. Are there other weighting factors that could maybe accurately correct for that or attempt to correct for that?

*Charlie Brown:* Charlie Brown from SRC. At the end of the day what you've got is a failure to match a list of names, addresses, date of births that you have with what you find in voting records. And so, one question is do you pay your respondents? ...

You do? Okay, no, because the reason I ask is because that means presumably you have an address that's good enough to get them to cash the check. And so now the probability that the problem lies with the address information that you have strikes me as just gone down dramatically. So, that's good.

And then, I guess, the other question is if you just looked at the voting records in general, how many of these could you find if you checked, say, postal records? So, another question is so you've got dates of birth that are 1900. You're never going to match those, but I'm kind of wondering, of the people in the voting records, what fraction of that data could, in principle, one match?

And so, I would exclude things that have implausible dates of birth, addresses that don't match any known address in the postal records, and if that's like 20 percent of the sample, well then it won't surprise me that you can't find 20 percent of your purported voters.

If that's 3.1 percent and you're telling me that you can't find 20 percent of your voters, I'm thinking, “Well, maybe I've got to think a little harder about what the problem is.” So, is there any just – forget about does it match your people, is there anything you can do about the quality of the name, address, date of birth information just in the voting records that would let us get any sense of whether it's realistic to expect to find a voter there?

*Matthew Berent:* Well, what Mike has found and what we've observed with our data is that there’s a lot of variability across the different states in terms of the errors that you can look at them and say, “Obviously, this information is wrong.” Okay? This person is not yet born, and they're voting. You think, “That's obviously wrong.” So –
Charlie Brown: Can I interrupt just one second? For a lot of voting purposes, like a one percent error rate is – that's disenfranchising a million people, you can't deal with that. For your purposes, one percent – big whoop. So, is this like ten percent of that? That's what I'm asking. Is that like ten percent of the sample? Or is that like one percent of the sample?

Matthew Berent: Well, we're talking about a value that's closer to – let's see, Skip, what was the number? Do you recall the number? Look at the report. I think it's close to five percent that have an error that we can say it's very unlikely that this information is accurate. Now that's five percent we're – based on our criteria – we're saying, “This doesn't look right.” That can include – that does not include – real errors that don't look like errors.

Jon Krosnick: Okay. Such optimistic findings. It's got – I mean, they are optimistic, right, about the survey data.